'Going Semantic'?

Information, Libraries & the 'Semantic Web'

The following contribution is a speculation about the future relation between semantic web technology (taking 'ontologies' as an example) and libraries – and this in turn implies some speculation regarding the future of the librarian profession in general in relation to the WWW information space.

However, in order to lay the ground for such speculations, an initial look at the evolution of the WWW and the reasons for the conception of the 'semantic web' as well as some conceptional clarification regarding key terms such as 'data', 'information' or 'knowledge will help to avoid philosophical pitfalls that are common when dealing with concepts such as semantic web 'ontologies'.

This framework then will enable me to make a few statements regarding the future potential of semantic web technology in a librarian context – or perhaps rather the future potential of librarian information services in the WWW information space.

The WWW before the 'semantic turn'

In order to understand the motivation for concepts such as the 'semantic web' one has to keep in mind some characteristics of the WWW information space that have turned out to be weak or insufficient from a scientific user perspective. The WWW initially was conceived as a mere data accumulation and transport machine based on the use of the HTTP-protocol and thus as a mere carrier medium with no assumptions regarding the content to be handled on this basis. As a result, a mere accumulative 'brute force' perspective has been prevailing for a long time creating impressive data dumps with very little genuine information value.

From an organizational point of view the WWW was (and largely still is) a strange blend of metaphors: a 'network' with 'sites', 'pages' and 'documents' largely without rules regarding identities and roles and thus not a very trustworthy place to go.

Information entropy (some kind of 'white noise') and lack of security thus were and still are two of the main weak spots of the WWW. Users are constantly struggling with them and this has stimulated defense reactions that until now have done very little to solve the actual problems but which simply cure some of the symptoms instead.

The reaction to the 'brute force' accumulation of data has been 'brute force' selection on the side of selection resulting in generations of search

engines, robots and other methods for data selection that have just one characteristic in common: they completely ignore the semantic value of the data they are processing.

The reaction to the problem of trust and security has been the multiplication of proprietary authentication and authorization schemes with a few players trying to impose their solution upon the rest (as in the struggle between the .net and Liberty Alliance communities) – and here again no appropriate solution has been found yet that would go beyond curing symptoms.

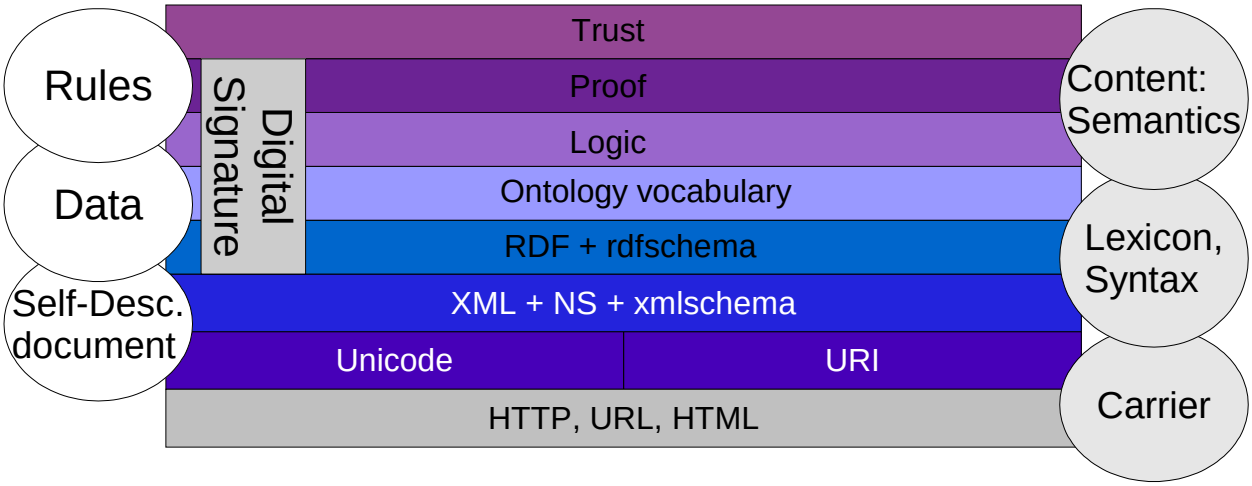
As a result, the WWW – in spite of its impressive amount of data – remains a very fragile information economy based on massive waste of resources needed to create the vital balance of accumulation and selection as well as for providing a minimum level of security, not to mention trust.

What is the 'Semantic Web'? Does it actually exist??

The 'semantic web' basically is an attempt to provide an appropriate remedy for both weaknesses. The assumption behind the term is that effective improvement of the WWW information economy would be possible only if machine methods could transcend the level of bits & bytes and were able to process semantics on top of that. It was assumed, that the „knowledge structures for the sake of machines“ needed for such a Semantic Web architecture could be obtained along the simple formula „Web + Knowledge Representation“ (Artur-Crofts-LeBoef2004).

shows the initial concept of the Semantic Web as postulated by its inventor, Tim Berners-Lee[1]:

1 Note that the stratification in the right (grey circles) is my addition! Berners-Lee 2001 is still a good choice for first reading and Davies2003 is a good introductions



Zeichnung 1: Semantic Web Layer Cake

From this image, it becomes clear that the two lower layers of the architecture still operate on carrier level only and that the following three levels up to and including ontologies basically operate on syntax level. Only the top three levels of logic, proof and trust are actually concerned with semantics in a genuine manner. In this overall picture, current stable semantic web technology ist available for the lower four levels and the ontology level is about to stabilize with the standardization af the Web Ontology Language (OWL). The 'semantic' Web thus still does not exist and remains a projection, but its foundations have become stable to some extent.

In order to give additional focus to this contribution I will therefore concentrate on semantic web ontologies in what follows.

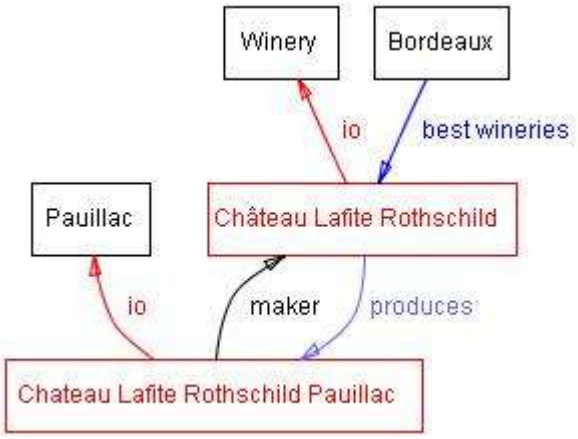
Semantic Web 'Ontologies'

Basic concepts ...

Noy2001gives a good definition of an ontology: „... an **ontology** is a formal explicit description of concepts in a domain of discourse (**classes** (sometimes called **concepts**)), properties of each concept describing various features and attributes of the concept (**slots** (sometimes called **roles** or **properties**)), and restrictions on slots (**facets** (sometimes called **role restrictions**)). An ontology together with a set of individual **instances** of classes constitutes a **knowledge base**. „

The constituent elements of such ontologies are triples containing subject, predicate and object (S->P->O), which usually are expressed using RDF. An example of such a triplet is the assertion 'speaker'->'HasName'->'Gradmann'.

An example of a small fragment of a knowledge base made up of such triplets again is given in Noy2001:



Drawing 2: A knowledge base fragment

These indications should be sufficient to give a first idea of what semantic web ontologies actually are. Elements and classes of such ontologies represent constituent elements of our 'world' on conceptional level using WWW technology such as RDF. Without going into much detail here, some potential misunderstandings and widespread confusions regarding semantics in the WWW context in general and 'ontologies' in particular should now become evident.

… Confusions …

Generally, when referring to networked bits and bytes we conceive these as 'data' or as 'information' or as 'knowledge' on different levels of thought and all too often implicitly confound these conceptional layers when using the term 'semantics'.

Furthermore, the very simplistic carrier-content model implicit in 'ontology'-conceptualization may well be fundamentally inappropriate from a semiological point of view, since the actual opposition one would need here is the one of signifier and significate, where both elements of the opposition are in turn twofold – however, this would profoundly harm the robust basic model of such 'ontologies'.

It does not help at all in this context that when conceiving 'content' in the WWW context most of us systematically confound 'things', pointers to 'things', meta-'things' and 'signs'.

But the main pitfall with regard to ontologies is a philosophical one and concerns the assumptions about 'truth' implicit in the use of the term! Benel2001 makes an important distinction in this regard with regard to different metaphysics of 'truth'. He opposes a *'positivist'* conception that

is based on a priori, that is consensual and assuming 'obvious' constituents of 'world' to the *conventionalist* and *hermeneutical* paradigms of 'truth'. It seems as if the very concept of 'ontologies' as outlined above only makes sense in the positivist paradigma and this entrails the question whether the other paradigms are de facto excluded from 'semantic web' knowledge management.

All these confusions have further blurred the concept of semantic web ontologies the name of which maybe has not been chosen very well right from the start. In order to clarify things it is therefore useful to point out that semantic web ontologies are by no means a panacea for all semantics based operations in the WWW – but that they can be very useful provided their limitations are well understood.

… and limitations

In general, semantic web ontologies probably are appropriate for modeling terms and concepts that are equivalent of those sections of our 'world', which we feel able to organize in taxonomic models of knowledge organization. They probably are not the best solution in all scenarios that are substantially built on 'interpretation' and which heavily depend on semiological complexity.

This means that 'evident' domains such as places, time periods or non-cultural artefacts such as cars or aeroplanes are good areas for the use of ontologies whereas for instance the semantics of cultural artefacts such as art and literature will be very difficult to model using such tools (if this is possible at all!)

In other words, ontologies are an appropriate means for all domains that can effectively be classified – but that does not mean that ontologies actually **are** classifications: they are part of a very specific architectural framework (cf. the layer model above) and they are conceived as a basis for automated operations on the semantics of WWW content.

'Ontologies' and Libraries

I have made a proposal elsewhere regarding the specific use that could be made of semantic web technology in the area of librariean metadata[2] and will therefore concentrate on the potential use of semantic web ontologies in the librarian context in this contribution.

As it seems, ontologies would fit in current librarian data models quite seamlessly as in the xml fragment reproduced below:

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE rdf:RDF [
```

2 Gradmann2005.

```
 [...]    xmlns:rdfs="&rdfs;">
<rdfs:Class rdf:ID="place">
   <rdfs:label xml:lang="en">place</rdfs:label>
   <rdfs:subClassOf rdf:resource="#work"/>
   <rdfs:subClassOf rdf:resource="#Corporate-Body"/>
</rdfs:Class>
[...]
</rdf:RDF>
```

The 'natural' place of some kind of ontology is indicated in the row printed in bold here – however, the actual **interest** anyone outside the librarian context might have in such an approach depends essentially on what mission 'Digital Libraries' will ultimately have and the place they will choose in the WWW information space.

If these 'Digital Libraries' will basically be self contained content stores using proprietary metadata standards and the WWW merely for data transfer - with a focus on *carrier* as regards the WWW - they will not need to be involved with 'ontologies' and SemWeb at all (but may face serious other problems because of such a choice!)

In case they are conceived as content stores integrated in WWW content architectures with WWW-transparent metadata standards – and thus with a focus on WWW *syntax* – a limited, yet systematic impact of semantic web technology and ontologies can be anticipated.

If 'Digital Libraries' were conceived as content stores and enablers for content based operations – with a clear focus on *semantics* thus – semantic web technology would even become core business of such 'libraries'; but this remains a very utopist and unlikely scenario ...

To conclude with I would therefore like to make a few pointed statements regarding the use of 'ontologies' in 'librarian contexts:

- 'Ontologies' have a potential of being useful in 'Digital Libraries' if these in turn are embedded in a WWW transparent information architectures.
- Using 'ontologies' will be a useless effort if this is part of the library automation paradigm as it is still basically valid today.
- The use of 'ontologies' can save a lot of human resources for intellectually demanding tasks - provided their limitations are well understood
- Ontologies will not be useful if regarded as a panacea for all problems of structuring, interoperability and reuse of librarian information models

- Nor will there be a appropriate use of 'ontologies' if this is conceived as continuing business (i. e. “classification”) just using other technical means - and thus probably not as part of 'Digital Library' settings in general.
- Finally, the use of 'Semantic Web' technology will be extremely limited if this means once again stepping in the traps of Artificial Intelligencel – and this statement evidently is not limited to the librarian environment.!

The majority of negative statements in this concluding section may lead towards two completely different overall conclusions: one of these would question the appropriateness of semantic web technology in the librarian context – the other would fundamentally question the way libraries have transposed their traditional functional paradigms to the WWW environment[3]. To determine which of these perspectives should be prevailing probably remains an act of faith for the time being ...

Selected Reading

Artur, Odile / Crofts, Nicholas / LeBoeuf, Patrick: Ontologies. Presentation at the ELAG Semantic Web and Libraries seminar. 19 April 2002 (available via http://www.ifnet.it/elag2002/programme.html) (=Artur2002)

Benel, Aurélien et al.: Truth in the Digital Library: From Ontological to Hermeneutical Systems. Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, September 4-9, 2001. Lecture Notes in Computer Science #2163. Heidelberg: Springer-Verlag 2001. pp.366-377. ISBN 3-540-42537-3 (=Benel2001)

Berners-Lee, Tim / Hendler, James / Lassila, Ora: The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In: Scientific American, 2001.5. (=Berners-Lee2001)

Gradmann, Stefan: How digital will Libraries ever be? Musings on the Limits of a Popular Metaphor. In: Informacné Správanie a Digitálne Kniznice / Information Behaviour in Digital Libraries. Bratislava 2003, pp. 27-40 (GRADMANN 2003) (=Gradmann2003)

Gradmann, Stefan: rdfs:frbr - Towards an implementation model for library catalogs using semantic web technology. In: Cataloguing and Classification Quarterly, FRBR Theme Issue, to be published Spring 2005 (=Gradmann2005)

3 Gradmann2003 is concerned with this issue.

Noy, Natalya F./McGuinness, Deborah: Ontology Development 101: A Guide to Creating Your First Ontology (http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html) (=Noy2001)

Towards the semantic web: ontology-driven knowledge management / edited by Jahn Davies, Dieter Fensel, and Frank van Harmelen. Chichester: Wiley 2003 (=Davies2003)